DØnote 5045-CONF

# A Study of $B_s^0$ Oscillations

The DØ Collaboration
URL *http://www-d0.fnal.gov*

(Dated: April 8, 2006)

A search for $B_s^0 - \bar{B}_s^0$ oscillations was performed with a large sample of semileptonic $B_s^0$ decays corresponding to approximately 1 fb$^{-1}$ of integrated luminosity accumulated with the DØ Detector in Run II at the Fermilab Tevatron. The flavor of the final state of the $B_s^0$ meson was determined using the muon charge from the partially reconstructed decay $B_s^0 \to D_s^- \mu^+ \nu X$, $D_s^- \to \phi\pi^-$, $\phi \to K^+K^-$. An opposite-side tagging method was used for the initial-state flavor determination. A log-likelihood scan was performed. The likelihood curve is well behaved and has a preferred value of the oscillation frequency $\Delta m_s = 19$ ps$^{-1}$, with a 90% confidence level interval of $17 < \Delta m_s < 21$ ps$^{-1}$, assuming Gaussian uncertainties. Ensemble tests indicate that if $\Delta m_s$ lies above our region of sensitivity ($> 22$ ps$^{-1}$), only 5.0% of the trial measurements give a measurement similar to or better than our observation anywhere in the window $16 < \Delta m_s < 22$ ps$^{-1}$.

*Preliminary Results for Winter 2006 Conferences*

## I. INTRODUCTION

At present, one of the most interesting topics in $B$ physics is the observation of $B_s^0$ oscillations and the measurement of the oscillation frequency $\Delta m_s$. These oscillations have yet to be observed experimentally and the current limit is $\Delta m_s > 16.6$ ps$^{-1}$ at the 95% confidence level [1]. A measurement of $\Delta m_s$ is an important test of the CKM formalism of the Standard Model, and combining it with a measurement of $\Delta m_d$, the oscillation frequency in the $B_d^0$ system, will allow us to reduce the uncertainty on the value of the CKM matrix element $V_{td}$. If the Standard Model is correct, and if the current experimental limits on $\Delta m_s$ are included in the fit, then $\Delta m_s = 18.3^{+6.5}_{-1.5}$ ps$^{-1}$ from global fits to the unitarity triangle. If information from $B_s^0$ oscillations limits is not included, global fits give $\Delta m_s = 20.9^{+4.5}_{-4.2}$ ps$^{-1}$ [2] or $\Delta m_s = 21.2 \pm 3.2$ ps$^{-1}$ [3].

## II. DETECTOR DESCRIPTION

The DØ detector is described in detail elsewhere [4, 5]. The following main elements of the DØ detector are essential for this analysis:

- The magnetic central-tracking system, which consists of a silicon microstrip tracker (SMT) and a central fiber tracker (CFT), both located within a 2-T superconducting solenoidal magnet;
- The liquid-argon/uranium calorimeter;
- The muon system located beyond the calorimeter.

The SMT has 800,000 individual strips, with typical pitch of $50 - 80$ $\mu$m, and a design optimized for tracking and vertexing capability at $|\eta| < 3$, where $\eta = -\ln(\tan(\theta/2))$ and $\theta$ is the polar angle. The CFT has eight thin coaxial barrels, each supporting two doublets of overlapping scintillating fibers of 0.835 mm diameter, one doublet being parallel to the collision axis, and the other alternating by $\pm 3^\circ$ relative to the axis. The resolution of the impact parameter with respect to the collision point is about 20 $\mu$m for 5 GeV/$c$ tracks.

The three components of the liquid-argon/uranium calorimeter are housed in separate cryostats. A central section, lying outside the tracking system, covers up to $|\eta| = 1.1$. Two end calorimeters extend the coverage to $|\eta| \approx 4$.

The muon system consists of a layer of tracking detectors and scintillation trigger counters inside a 1.8 T iron toroid, followed by two additional layers outside the toroid. Tracking at $|\eta| < 1$ relies on 10 cm wide drift tubes, while 1-cm mini-drift tubes are used at $1 < |\eta| < 2$.

## III. DATA SAMPLE

This analysis used a $B_s^0 \to D_s^- \mu^+ \nu X$, $D_s^- \to \phi\pi^-$, $\phi \to K^+K^-$ data sample selected with an offline filter from all data taken from April 2002 to January 2006 with no explicit trigger requirement, although most of the sample was collected by single muon triggers. The selections for the offline filter are described below. Charge conjugate states are assumed throughout this paper.

For this analysis, the muons were required to have $p_T > 2$ GeV/$c$ and $p > 3$ GeV/$c$, to have at least one hit each in the CFT and SMT, and to have measurements in at least two layers of the muon chambers.

The primary vertex position in the transverse plane was determined on an event-by-event basis by requiring the tracks in the event to come from a common collision point that is constrained by the mean beam-spot position calculated on a run-by-run basis. The tracks used in the reconstruction of the $B_s^0$ semileptonic decay were excluded from the primary vertex fit.

All charged particles in the event were clustered into jets using the DURHAM clustering algorithm [6] with a $p_T$ cut-off parameter set at 15 GeV/$c$ [7]. The $D_s^-$ candidate was constructed from three tracks included in the same jet as the reconstructed muon. Two oppositely charged tracks were assigned the kaon mass and were required to form a $\phi$ meson satisfying $1.004 < M(K^+K^-) < 1.034$ GeV/$c^2$. The third track was assigned the pion mass and was required to have a charge opposite to that of the muon. All three tracks were required to have hits in the SMT and CFT. The transverse momentum requirements were $p_T > 0.7$ GeV/$c$ for the kaons and $p_T > 0.5$ GeV/$c$ for the pion. The three tracks were required to form a common $D_s^-$ vertex with $\chi^2_D < 16$ for the vertex fit. The vertexing algorithm is described in detail in Ref. [8]. For each particle, the transverse $\epsilon_T$ and longitudinal $\epsilon_L$ projections of the track impact parameter with respect to the primary vertex, together with the corresponding uncertainties $\sigma(\epsilon_T)$ and $\sigma(\epsilon_L)$, were computed. The combined significance $(\epsilon_T/\sigma(\epsilon_T))^2 + (\epsilon_L/\sigma(\epsilon_L))^2$ was required to be greater than 4 for the kaons. The distance $d_T^D$ between the primary and $D_s^-$ vertices in the transverse plane was required to exceed 4

standard deviations, that is, $d_T^D/\sigma(d_T^D) > 4$. The angle $\alpha_T^D$ between the $D_s^-$ momentum and the direction from the primary vertex to the $D_s^-$ vertex in the transverse plane was required to satisfy the condition $\cos(\alpha_T^D) > 0.9$.

The tracks of the muon and $D_s^-$ candidate were required to produce a common $B_s^0$ vertex with $\chi^2_B < 9$ for the vertex fit. The transverse momentum of the $B_s^0$ candidate, $p_T^{\mu^+ D_s^-}$, was defined as the vector sum of the transverse momenta of the muon and the $D_s^-$ candidate. The mass of the $(\mu^+ D_s^-)$ system was required to be within the range $2.6 < M(\mu^+ D_s^-) < 5.4$ GeV/$c^2$. The transverse decay length of the $B_s^0$ hadron, $d_T^B$, was defined as the distance in the transverse plane between the primary vertex and the vertex produced by the muon and the $D_s^-$ meson. If the distance $d_T^B$ exceeded $4 \cdot \sigma(d_T^B)$, the angle $\alpha_T^B$ between the $B_s^0$ momentum and the direction from the primary to the $B_s^0$ vertex in the transverse plane was required to satisfy the condition $\cos(\alpha_T^B) > 0.95$. The distance $d_T^B$ was allowed to be greater than $d_T^D$, provided that the distance between the $B_s^0$ and $D_s^-$ vertices, $d_T^{BD}$, was less than $2 \cdot \sigma(d_T^{BD})$.

The final event samples were selected using a Likelihood Ratio Method, described below.

### A. Likelihood Ratio Method

In the Likelihood Ratio Method, a set of discriminating variables, $x_1, ...x_n$, is constructed for a given event. The probability density functions (*pdf*s), $f_i^s(x_i)$ for the signal and $f_i^b(x_i)$ for the background, are built for each variable $x_i$. The combined selection variable $y$ is defined as

$$y = \prod_{i=1}^{n} y_i; \quad y_i = \frac{f_i^b(x_i)}{f_i^s(x_i)}. \tag{1}$$

The variable $x_i$ can be undefined for some events. In this case, the corresponding variable $y_i$ is set to unity. The selection of the signal is obtained by applying the cut $y < y_0$ [9]. For uncorrelated variables $x_1, ...x_n$, the selection using the combined variable $y$ gives the best possible performance, i.e., the maximal signal efficiency for a given background efficiency.

The following discriminating variables were used:

- Helicity angle, defined as the angle between the $D_s^-$ and $K^+$ momenta in the $(K^+K^-)$ center-of-mass system. (The $K^+$ and $K^-$ mesons decay back-to-back in the $\phi$ rest frame so the choice of $K^+$ over $K^-$ is arbitrary);
- Isolation, computed as $Iso = p^{tot}(\mu^+ D_s^-)/(p^{tot}(\mu^+ D_s^-) + \sum p_i^{tot})$. The sum $\sum p_i^{tot}$ is taken over all charged particles in the cone $\sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < 0.5$, where $\Delta\eta$ and $\Delta\phi$ are the pseudorapidity and the azimuthal angle with respect to the $(\mu^+ D_s^-)$ direction. The $\mu^+$, $K^+$, $K^-$ and $\pi^-$ momenta are not included in the sum;
- $p_T(K^+K^-)$;
- Invariant mass, $M(\mu^+ D_s^-)$;
- $\chi^2$ of the $D_s^-$ vertex fit;
- Invariant mass, $M(K^+K^-)$.

The probability density functions were constructed using the real data events. For each channel, three bands, $B_1$, $B_2$ and $S$, were defined as:

$$\begin{aligned} B_1 &: 1.75 < M(D_s^-) < 1.79 \text{ GeV}/c^2; \\ B_2 &: 2.13 < M(D_s^-) < 2.17 \text{ GeV}/c^2; \\ S &: 1.92 < M(D_s^-) < 2.00 \text{ GeV}/c^2. \end{aligned}$$

The background probability density function for each variable was constructed using events from the $B_1$ and $B_2$ bands. The signal probability density function was constructed by subtracting the background, obtained as a sum of the distributions in the $B_1$ and $B_2$ bands, from the distribution of events in band $S$.

The final cut on the combined variable, $-\log_{10} y > 0.12$, was selected by requiring the maximal value of $N_S/\sqrt{N_S + N_{B_1} + N_{B_2}}$, where $N_S$, $N_{B_1}$ and $N_{B_2}$ are the number of events in bands $S$, $B_1$ and $B_2$, respectively. The total number of $D_s^-$ candidates passing this combined variable cut in the mass peak is $26,710 \pm 556$ (stat.), while the number of $D^-$ candidates is $7,422 \pm 281$ (stat.) (Fig. 1). There are $5,601 \pm 102$ (stat). $B_s^0$ candidates that have an identified initial-state flavor obtained from the opposite-side tag (Fig. 2), as explained below.

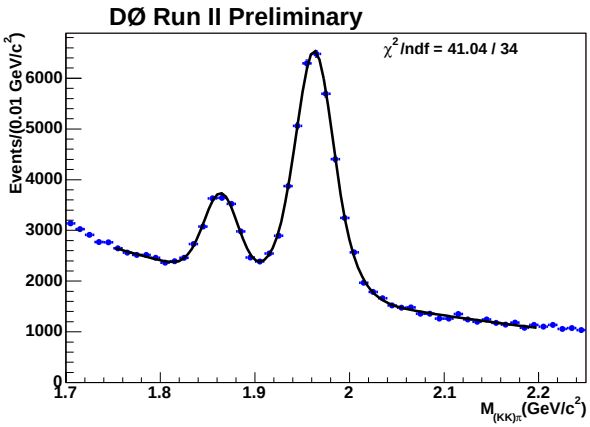


FIG. 1: $M(K^+K^-\pi^-)$ invariant mass distribution for the untagged $B_s^0$ sample. The left and right peaks correspond to $\mu^+D^-$ and $\mu^+D_s^-$ candidates, respectively. The curve represents the fit function to this mass spectrum. For fitting the mass spectra, a single Gaussian was used to describe the $D^- \to \phi\pi^-$ decays and a double Gaussian was used for the $D_s^- \to \phi\pi^-$ decays. The background is modeled by an exponential.

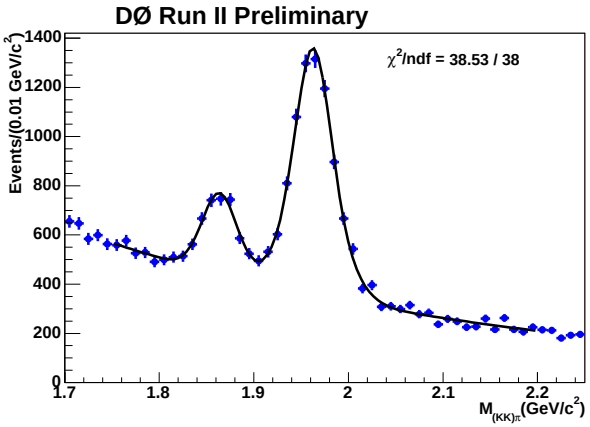


FIG. 2: $M(K^+K^-\pi^-)$ invariant mass distribution for the tagged $B_s^0$ sample. The left and right peaks correspond to $\mu^+D^-$ and $\mu^+D_s^-$ candidates, respectively. The curve represents the fit function to this mass spectrum. For fitting the mass spectra, a single Gaussian was used to describe the $D^- \to \phi\pi^-$ decays and a double Gaussian was used for the $D_s^- \to \phi\pi^-$ decays. The background is modeled by an exponential.

## IV. FLAVOR TAGGING

A necessary step in the $B_s^0$ oscillation analysis is the determination of the $B_s^0/\bar{B}_s^0$ initial- and final-state flavors. The presence of the muon in the $B_s^0$ semileptonic decay allows a determination of the final-state flavor since the $b$-quark flavor is correlated with the charge of the muon in the decays $B_s^0 \to \mu^+X$ and $\bar{B}_s^0 \to \mu^-X$.

The opposite-side tagging (OST) [10] of the initial flavor of the $B_s^0$ meson exploits the fact that in $b\bar{b}$ pair production two $b$-flavored hadrons are always produced. Partial reconstruction of the $b$ hadron on the opposite side to the $B_s^0$ gives information on the initial flavor of the $B_s^0$. Purity, dilution and tagging efficiency are three important parameters for describing the tagging performance. The purity of the tagging method is defined as $\eta_s = N_{correctly\ tagged\ events}/N_{total\ tagged\ events}$. The dilution is related to the purity by the simple formula $\mathcal{D} = 2\eta_s - 1$. Finally, the tagging efficiency is defined as $\epsilon = N_{total\ tagged\ events}/N_{total\ events}$. A Likelihood Ratio Method was used again, where a set of flavor discriminating variables, $x_1, ...x_n$, was constructed for each event. In the construction of these variables, an object, either a lepton $\ell$ (electron or muon) or a reconstructed secondary vertex (SV), was defined to be on the opposite side from the $B_s^0$ meson if it satisfied $\cos\phi(\vec{p}_{\ell \text{ or SV}}, \vec{p}_B) < 0.8$, where $\vec{p}_B$ is the reconstructed three-momentum of the $B_s^0$ meson, and $\phi$ is the azimuthal angle. A lepton jet charge was formed as $Q_J^\ell = \sum_i q^i p_T^i / \sum_i p_T^i$, where the sum is over all charged particles, including the lepton but excluding the $B_s^0$ decay products, inside a cone of $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < 0.5$ centered on the lepton.

Another discriminating variable is the secondary vertex charge, defined as $Q_{\rm SV} = \sum_i (q^i p_L^i)^{0.6} / \sum_i (p_L^i)^{0.6}$, where the sum is over all charged particles associated with the secondary vertex, and $p_L^i$ is the longitudinal momentum of track $i$ with respect to the direction of the secondary vertex momentum. Finally, event charge is defined as $Q_{\rm EV} = \sum_i q^i p_T^i / \sum_i p_T^i$, where the sum is over all tracks with $p_T > 0.5$ GeV outside a cone of $\Delta R > 1.5$ centered on the $B_s^0$ direction.

For an initial $b$ ($\bar{b}$) quark, the *pdf* for a given variable $x_i$ is denoted $f^b(x_i)$ ($f^{\bar{b}}(x_i)$), and the combined tagging variable $d_{\rm tag}$ is defined as

$$d_{\rm tag} = \frac{1-z}{1+z}; \qquad z = \prod_{i=1}^{n} \frac{f_i^{\bar{b}}(x_i)}{f_i^b(x_i)}. \tag{2}$$

The variable $d_{\rm tag}$ varies between $-1$ and 1. An event with $d_{\rm tag} < 0$ ($> 0$) was tagged as a $b$ ($\bar{b}$) quark, and $|d_{\rm tag}|$ is related to $\mathcal{D}$, with larger $|d_{\rm tag}|$ values corresponding to higher purity. The *pdf* of each discriminating variable was found in data for $b$ and $\bar{b}$ quarks using a large sample of $B^+ \to \mu^+\nu D^0$ events where the initial state is known from the charge of the decay muon.

A subsample of all $B$ candidates for which the partial reconstruction on the opposite side was available is called the "*total tagged events*" sample. $B_d^0$ mesons oscillate with low frequency while $B^+$ mesons do not oscillate. Therefore,

the $B_d^0$ and $B^+$ samples are used to determine the number of "*correctly tagged events*" and, therefore, to calibrate the OST.

Ref. [10] describes a measurement of the $B_d^0$ oscillation frequency and a determination of the dilution for the $B_d^0$ and $B^+$ samples. Each tagged $B$ candidate is characterized by a variable $d_{pr}$, which gives a prediction of the dilution for that candidate using the formulas

$$\mathcal{D}(d_{pr})\,\big|_{d_{pr}<0.6} = 0.457\cdot|d_{pr}| + 2.349\cdot|d_{pr}|^2 - 2.498\cdot|d_{pr}|^3, \qquad (3)$$
$$\mathcal{D}(d_{pr})\,\big|_{d_{pr}>0.6} = 0.6.$$

Another parameterization, $\mathcal{D}(d_{pr})$, was used to estimate the systematic uncertainty:

$$\mathcal{D}(d_{pr}) = \frac{0.6}{1+\exp\left(-\frac{d_{pr}-0.312}{0.108}\right)}. \qquad (4)$$

## V. EXPERIMENTAL OBSERVABLES

The proper lifetime of the $B_s^0$ meson, $c\tau_{B_s^0}$, for semileptonic decays can be written as

$$c\tau_{B_s^0} = x^M\cdot K, \quad \text{where} \quad x^M = \left[\frac{\mathbf{d_T^B}\cdot\mathbf{p_T^{\mu D_s^-}}}{(p_T^{\mu D_s^-})^2}\right]\cdot cM_B. \qquad (5)$$

$x^M$ is the *visible proper decay length*, or VPDL, and $K$ is the correction factor, also called the K factor. Semileptonic $B$ decays necessarily have an undetected neutrino present in the decay chain, making a precise determination of the kinematics for the $B$ meson impossible. In addition, other neutral or non-reconstructed charged particles can be present in the decay chain of the $B$ meson. This leads to a bias towards smaller values of the $B$ momentum, which is calculated using the reconstructed particles. A common practice to correct this bias is to scale the measured momentum of the $B$ candidate by a $K$ factor, which takes into account the effects of the neutrino and other lost or non-reconstructed particles. For this analysis, the $K$ factor was defined as

$$K = p_T(\mu^+ D_s^-)/p_T(B_s^0), \qquad (6)$$

where $p_T$ denotes the absolute value of the transverse momentum. The $K$-factor distributions used to correct the data were obtained from the Monte Carlo (MC) simulation.

## VI. FITTING PROCEDURE

All tagged events with $1.72 < M(K^+K^-\pi^-) < 2.22$ GeV/$c^2$ were used in the unbinned likelihood fitting procedure. The likelihood for an event to arise from a specific source in the sample depends on the $x^M$, its uncertainty ($\sigma_{x^M}$), the mass of the $D_s^-$ meson candidate ($m$), the predicted dilution ($d_{pr}$) and the selection variable $y$ described in Section III A. All of the quantities used in the unbinned likelihood fitting procedure are known on an event-by-event basis. The *pdf* for each source can be expressed by the product of the corresponding *pdf*s:

$$f_i = P_i^{x^M}(x^M, \sigma_{x^M}, d_{pr}) P_i^{\sigma_{x^M}} P_i^m P_i^{d_{pr}} P_i^y. \qquad (7)$$

The VPDL *pdf* $P_i^{x^M}(x^M, \sigma_{x^M}, d_{pr})$ represents a conditional probability, therefore it should be multiplied by $P_i^{\sigma_{x^M}}$ and $P_i^{d_{pr}}$ to have a joint *pdf* (see "Probability" section in PDG [12]). The *pdf*s $P_i^m$ and $P_i^y$ are used for separation of signal and background.

The following sources, $i$, were considered:

- $\mu^+ D_s^-(\to \phi\pi^-)$ signal with fraction $\mathcal{F}_{\mu D_s}$.
- $\mu^+ D^-(\to \phi\pi^-)$ signal with fraction $\mathcal{F}_{\mu D^\pm}$.
- $\mu^+ D^-(\to K\pi\pi^-)$ reflection with fraction $\mathcal{F}_{\mu D^\pm_{refl}}$. The reflection arises due to mass misassignment in this channel. The $D^-$ mass peak shifts to $\sim 2$ GeV/$c^2$ if the kaon mass is incorrectly assigned to one of the pion tracks.

- Combinatorial background with fraction $(1 - \mathcal{F}_{\mu D_s} - \mathcal{F}_{\mu D^\pm} - \mathcal{F}_{\mu D^\pm_{refl}})$.

The fractions $\mathcal{F}_{\mu D_s}$ and $\mathcal{F}_{\mu D^\pm}$ were determined from the mass fit (see Fig. 2). The total probability density function for a $B$ candidate has the form

$$F_n = \mathcal{F}_{\mu D_s} f_{\mu D_s} + \mathcal{F}_{\mu D^\pm} f_{\mu D^\pm} + \mathcal{F}_{\mu D^\pm_{refl}} f_{\mu D^\pm_{refl}} + \left(1 - \mathcal{F}_{\mu D_s} - \mathcal{F}_{\mu D^\pm} - \mathcal{F}_{\mu D^\pm_{refl}}\right) f_{bkg}. \tag{8}$$

The following form was minimized using the MINUIT [11] program:

$$\mathcal{L} = -2 \sum_n \ln F_n, \tag{9}$$

where $n$ varies from 1 to $N_{total\ tagged\ events}$.

The *pdf*s for the VPDL uncertainty ($P_i^{\sigma_{x^M}}$), mass ($P_i^m$), dilution ($P_i^{d_{pr}}$), and selection variable $y$ ($P_i^y$) were taken from experimental data. The signal *pdf*s were also used for the $\mu^+ D^-(\to \phi\pi^-)$ signal and the $\mu^+ D^-(\to K^+\pi^-\pi^-)$ reflection. The dependence of the background slope on VPDL was also taken into account. The mass *pdf* for the $\mu^+ D^-(\to K^+\pi^-\pi^-)$ reflection was determined from the MC. The fraction $\mathcal{F}_{\mu D^-_{refl}}$ of $K^+\pi^-\pi^-$ reflected events under the $K^+K^-\pi^-$ curve was determined using a fit to the full untagged $M(K^+K^-\pi^-)$ mass spectrum (Fig. 1) and was found to be less than 1% of the number of signal $\mu^+ D_s^-(\to \phi\pi^-)$ events.

### A. *pdf* for $\mu^+ D_s^-$ Signal

The $\mu^+ D_s^-$ sample is composed mostly of $B_s^0$ mesons with some contributions from $B^+$ and $B_d^0$ mesons. Different species of $B$ mesons behave differently with respect to oscillations. Neutral $B_d^0$ and $B_s^0$ mesons do oscillate (with different frequencies) while charged $B^+$ mesons do not. The possible contributions of $b$ baryons to the sample were estimated to be small and so are neglected.

The data sample is divided into non-oscillated and oscillated subsamples as determined by the flavor tagging. For a given type of $B_q$ hadron, where $q$ is $d$, $u$, or $s$, the distribution of the visible proper decay length $x$ for non-oscillated and oscillated cases ($p^{nos}$ and $p^{osc}$) is given by:

$$p_s^{nos}(x, K, d_{pr}) = \frac{K}{c\tau_{B_s^0}} \exp(-\frac{Kx}{c\tau_{B_s^0}}) \cdot 0.5 \cdot (1 + \mathcal{D}(d_{pr}) \cos(\Delta m_s \cdot Kx/c)) \tag{10}$$

$$p_s^{osc}(x, K, d_{pr}) = \frac{K}{c\tau_{B_s^0}} \exp(-\frac{Kx}{c\tau_{B_s^0}}) \cdot 0.5 \cdot (1 - \mathcal{D}(d_{pr}) \cos(\Delta m_s \cdot Kx/c)) \tag{11}$$

$$p_{DsDs}^{osc}(x, K) = \frac{K}{c\tau_{B_s^0}} \exp(-\frac{Kx}{c\tau_{B_s^0}}) \cdot 0.5 \tag{12}$$

$$p_{DsDs}^{nos}(x, K) = \frac{K}{c\tau_{B_s^0}} \exp(-\frac{Kx}{c\tau_{B_s^0}}) \cdot 0.5 \tag{13}$$

$$p_u^{nos}(x, K, d_{pr}) = \frac{K}{c\tau_{B^+}} \exp(-\frac{Kx}{c\tau_{B^+}}) \cdot 0.5 \cdot (1 - \mathcal{D}(d_{pr})) \tag{14}$$

$$p_u^{osc}(x, K, d_{pr}) = \frac{K}{c\tau_{B^+}} \exp(-\frac{Kx}{c\tau_{B^+}}) \cdot 0.5 \cdot (1 + \mathcal{D}(d_{pr})) \tag{15}$$

$$p_d^{nos}(x, K, d_{pr}) = \frac{K}{c\tau_{B_d^0}} \exp(-\frac{Kx}{c\tau_{B_d^0}}) \cdot 0.5 \cdot (1 - \mathcal{D}(d_{pr}) \cos(\Delta m_d \cdot Kx/c)) \tag{16}$$

$$p_d^{osc}(x, K, d_{pr}) = \frac{K}{c\tau_{B_d^0}} \exp(-\frac{Kx}{c\tau_{B_d^0}}) \cdot 0.5 \cdot (1 + \mathcal{D}(d_{pr}) \cos(\Delta m_d \cdot Kx/c)). \tag{17}$$

Here $\tau_{B_q}$ is the lifetime of the $B_q$ hadron, where $q$ is $u, d$, or $s$. Note that there is a sign swap in Eqs. 14–17 with respect to Eqs. 10 and 11 due to anti-correlation of charge for muons from $B \to DD_s^-$; $D \to \mu^+ X$ processes.

The translation from real VPDL, $x$, to the measured VPDL, $x^M$, is achieved by a convolution of the $K$ factors and resolution functions as specified below.

$$P_j^{osc,\ nos}(x^M, \sigma_{x^M}, d_{pr}) \ = \tag{18}$$
$$\int_{K_{min}}^{K_{max}} dK\ D_j(K) \cdot \frac{Eff_j(x^M)}{N_j(K, \sigma_{x^M}, d_{pr})} \qquad \int_0^\infty dx\ G(x - x^M, \sigma_{x^M}) \cdot p_j^{osc,\ nos}(x, K, d_{pr}).$$

Here

$$G(x - x^M, \sigma_{x^M}) = \frac{1}{\sqrt{2\pi}\sigma_{x^M}} \exp\left(-\frac{(x - x^M)^2}{2\sigma_{x^M}^2}\right) \tag{19}$$

is the detector resolution of the VPDL and $Eff_j(x)$ is the reconstruction efficiency for a given decay channel $j$ of this type of $B$ meson as a function of VPDL. The function $D_j(K)$ gives the normalized distribution of the $K$ factor in a given channel $j$. The normalization factor $N_j$ is calculated by integration over the entire VPDL region:

$$N_j(K, \sigma_{x^M}, d_{pr}) \ = \tag{20}$$
$$\int_{-\infty}^{\infty} dx^M\ Eff_j(x^M) \ \cdot \ \int_0^\infty dx\ G(x - x^M, \sigma_{x^M}) \cdot \left(p_j^{osc}(x, K, d_{pr}) + p_j^{nos}(x, K, d_{pr})\right)$$

The total VPDL *pdf* for the $\mu^+ D_s^-$ signal is a sum of all the contributions that yield the $D_s^-$ mass peak:

$$P_{\mu D_s}^{osc,\ nos}(x^M, \sigma_{x^M}, d_{pr}) = (1 - \mathcal{F}_{peak}) \sum_j Br_j \cdot P_j^{osc,\ nos}(x^M, \sigma_{x^M}, d_{pr}) + \mathcal{F}_{peak} \cdot P_{peak}^{osc,\ nos}(x^M) \tag{21}$$

Here the sum $\sum_j$ is taken over all decay channels that yield a $\mu^+ D_s^-$ final state and $Br_j$ is the branching rate of a given channel $j$. In addition to the long-lived $\mu^+ D_s^-$ candidates from $B$ meson decays, there is a contribution, with fraction $\mathcal{F}_{peak}$, of the "peaking background", which consists of combinations of $D_s^-$ mesons and muons originating from different $c$ or $b$ quarks. The direct $c$ production gives the largest contribution to this background and, therefore, the function $P_{peak}^{osc,\ nos}(x^M)$ was determined from $c\bar{c}$ MC. We assume that this background produces negative and positive flavor tags with equal probability.

The choice of oscillated or non-oscillated VPDL *pdf* for Eq. 7 is made using relative charge of the muon from the $B_s^0$ meson with respect to the sign of $d_{pr}$:

$$d_{pr} \cdot q_\mu > 0: \ P^{x^M}(x^M, \sigma_{x^M}, d_{pr}) = P_{\mu D_s}^{osc}(x^M, \sigma_{x^M}, d_{pr}), \tag{22}$$
$$d_{pr} \cdot q_\mu < 0: \ P^{x^M}(x^M, \sigma_{x^M}, d_{pr}) = P_{\mu D_s}^{nos}(x^M, \sigma_{x^M}, d_{pr}).$$

The branching rates $Br_j$ were taken from the PDG [12]. The functions $D_j(K)$ and $Eff_j(x)$ were taken from the MC simulation, as explained later. The lifetimes of the $B^+$ and $B_d^0$ mesons were taken from PDG while the $B_s^0$ lifetime was measured using the total tagged $\mu^+ D_s^-$ sample.

### B. *pdf* for Combinatorial Background

The following contributions to the combinatorial background were considered:

1. Prompt background, with *pdf* $P_{bkg}$ and with the $\mu^+ D_s^-$ vertex coinciding with the primary vertex (described as a Gaussian with a width determined by the resolution; fraction in the background: $\mathcal{F}_0$). The resolution scale factor for this background is different from the signal resolution scale factor. The scale factor is a free fit parameter, $s_{bkg}$.

2. Background (*pdf* $P_{bkg}^{res}$) with quasi-vertices distributed around the primary vertex (described as a Gaussian with constant width $\sigma_{peak\_bkg}$; fraction in the background: $\mathcal{F}_{peak\_bkg}$).

3. Long-lived background, with *pdf* $p_{bkg}^{long}$ (exponential with constant decay length $c\tau_{bkg}$ convoluted with the resolution). This background was divided into three subsamples:

(a) insensitive to the tagging (fraction in the long-lived background: $(1-\mathcal{F}_{tsens})$);

(b) sensitive to the tagging and non-oscillating (fraction in the background sensitive to the tagging: $(1-\mathcal{F}_{osc})$);

(c) sensitive to the tagging and oscillating with frequency $\Delta m_d$ (fraction in the background sensitive to the tagging: $\mathcal{F}_{osc}$).

The fractions of these contributions and their parameters were determined from the data sample. The background *pdf* was expressed in the following form:

$$
\begin{aligned}
P_{bkg}(x^M,\sigma_{x^M},d_{pr}) &= \mathcal{F}_{peak\_bkg}G(0-x^M,\sigma_{peak\_bkg})+(1-\mathcal{F}_{peak\_bkg})\cdot P^{res}_{bkg}(x^M,\sigma_{x^M}), \qquad (23)\\
P^{res}_{bkg}(x^M,\sigma_{x^M},d_{pr}) &= \frac{Eff(x^M)}{N}\int_0^\infty dx\left(\mathcal{F}_0G(x-x^M,s_{bkg}\sigma_{x^M})\delta(x)+(1-\mathcal{F}_0)G(x-x^M,\sigma_{x^M})\cdot p^{long}_{bkg}\right),\\
p^{long,osc/nos}_{bkg}(x,d_{pr}) &= \frac{1}{c\tau_{bkg}}\exp\left(-\frac{x}{c\tau_{bkg}}\right)\left((1-\mathcal{F}_{tsens})+\mathcal{F}_{tsens}\left((1\pm\mathcal{D})(1-\mathcal{F}_{osc})+(1\pm\mathcal{D}\cos\left(\Delta m_d\cdot x/c\right))\cdot\mathcal{F}_{osc}\right)\right),
\end{aligned}
$$

where $N$ is a normalization constant and the fit parameters are $\mathcal{F}_{peak\_bkg}$, $\sigma_{peak\_bkg}$, $\mathcal{F}_0$, $\mathcal{F}_{tsens}$, $\mathcal{F}_{osc}$ and $c\tau_{bkg}$. As an efficiency $Eff(x^M)$, the efficiency for the $B^0_d\to D^-\mu^+\nu X$ channel was used.

## VII. FIT INPUTS

We have used the following measured parameters for $B$ mesons from the PDG [12] as inputs for the lifetime fitting procedure: $c\tau_{B^+}=501$ $\mu$m, $c\tau_{B^0_d}=460$ $\mu$m, and $\Delta m_d=0.502$ ps$^{-1}$. The latest PDG values were also used to determine the branching fractions of decays contributing to the $D^-_s$ sample. We used the event generator EvtGen [14] since this code was developed specifically for the simulation of $B$ decays. For those branching fractions not given in the PDG, we used the values provided by EvtGen, which are motivated by theoretical considerations. Taking into account the corresponding branching rates and reconstruction efficiencies, we calculated the contributions to our signal region from the various processes. The $B^0_s\to D^-_s\mu^+\nu X$ modes (including through $D^{*-}_s$, $D^{*-}_{s0}$, and $D^{'-}_{s1}$ and $\mu^+$ originating from $\tau$ decays) comprise $85.6\pm3.3\%$ of our sample, including reconstruction efficiency. Other backgrounds with both a real $D^-_s$ and $\mu^+$ and showing up in the peak, but not expected to oscillate with $\Delta m_s$, that are considered are $B\to D^+_{(s)}D^-_sX$ decays followed by $D^+_{(s)}\to\mu^+\nu X$. The assigned uncertainty to each channel covers possible trigger efficiency biases. We then determined the efficiency of the lifetime selections for the sample as a function of VPDL, as shown in Fig. 3 for the decay $B^0_s\to D^-_s\mu^+\nu X$.

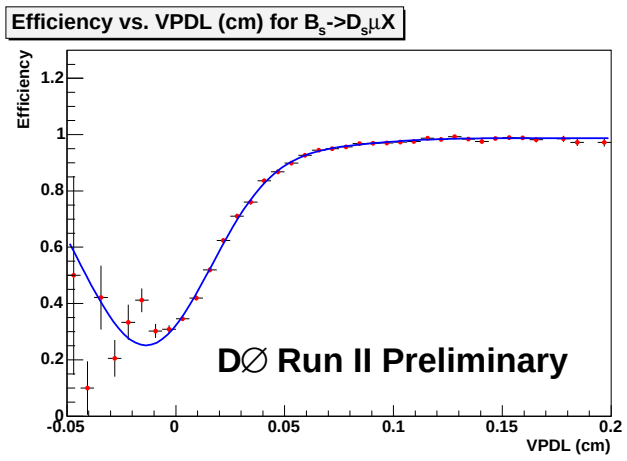


FIG. 3: Efficiency of the lifetime-dependent cuts as a function of VPDL for $B^0_s\to D^-_s\mu^+\nu X$.

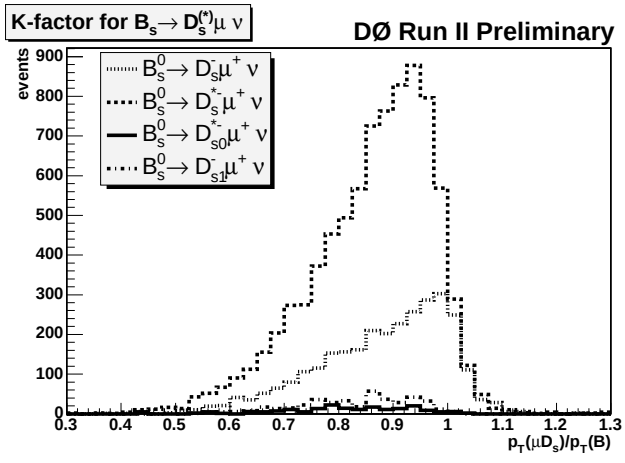


FIG. 4: $K$ factor distributions for $B^0_s\to\mu^+\nu D^-_s$; $B^0_s\to\mu^+\nu D^{-*}_s\to\mu^+\nu D^-_s$; $B^0_s\to\mu^+\nu D^{*-}_{s0}\to\mu^+\nu D^-_s$; $B^0_s\to\mu^+\nu D^{'-}_{s1}\to\mu^+\nu D^-_s$ processes.

In determining the $K$ factor distributions, MC generator-level information was used for the computation of $p_T$. Following the definition used in Eq. 6, the $K$ factor distributions for all considered decays were determined. Figure 4 shows the distributions of the $K$ factors for the semi-muonic decays of the $B^0_s$ meson. As expected, the $K$ factors for

$D_s^{-*}$, $D_{s0}^{*-}$ and $D_{s1}^{'-}$ have lower mean values because more decay products are lost. Note that since the $K$ factors in Eq. 6 were defined as the ratio of transverse momenta, they can exceed unity.

The VPDL uncertainty was estimated by the vertex fitting procedure. A resolution scale factor was introduced to take into account a possible bias. It was determined using the $J/\psi$ sample. Figure 5 shows the pull distribution, $PDL_{J/\psi}/\sigma(PDL_{J/\psi})$, of the $J/\psi$ vertex position with respect to that of the primary vertex, where $PDL$ is the *proper decay length*. The negative tail of the pull distribution of the $J/\psi$ vertex position with respect to that of the primary vertex should be a Gaussian with a sigma of unity if uncertainties assigned to the vertex coordinates are correct. We ignore the positive side of the pull distribution as that tends to be biased towards larger values due to $J/\psi$ mesons from real $B$ meson decays. For this study we exclude muons from $J/\psi$ decays from the primary vertex. The resulting pull distribution was fitted using a double Gaussian: the narrow Gaussian with width $\sigma_{narrow} = 0.998$ comprises 72% of the events, and the wide Gaussian with width $\sigma_{wide} = 1.775$ comprises 28%.

It is known that the scale factor depends on track transverse momenta. We took this dependence into account using the scale factor determined for $J/\psi$ candidates where the leading muon has $p_{t_\mu} > 6$ GeV/$c$ to estimate a contribution to the systematic uncertainty. The corresponding scale factors increased by 2.5%.

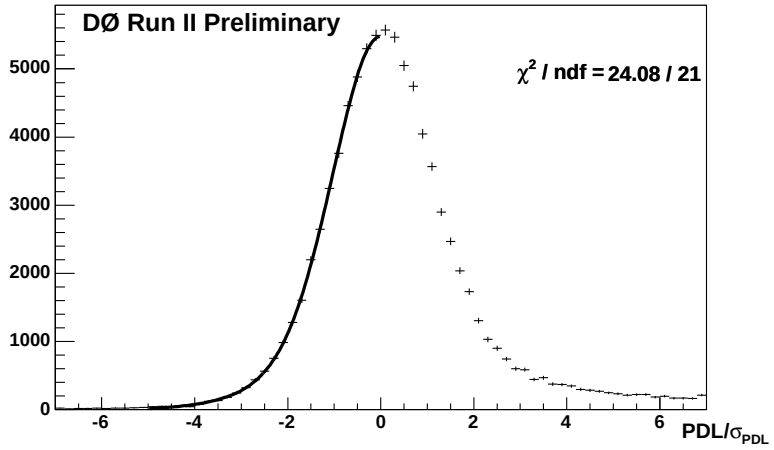


FIG. 5: Pull distribution of the $J/\psi$ vertex position with respect to that of the primary vertex: $PDL_{J/\psi}/\sigma(PDL_{J/\psi})$.

The total tagged data sample was used to determine the parameters: $\mathcal{F}_{peak\_bkg} = 0.022 \pm 0.007$, $\sigma_{peak\_bkg} = 117 \pm 11$ $\mu$m, $\mathcal{F}_0 = 0.127 \pm 0.008$, $s_{bkg} = 2.14 \pm 0.03$, $c\tau_{bkg} = 627 \pm 6$ $\mu$m, $\mathcal{F}_{peak} = 0.027 \pm 0.006$, $\mathcal{F}_{tsens} = 0.664 \pm 0.044$, $\mathcal{F}_{osc} = 0.512 \pm 0.053$ and $c\tau_{B_s^0} = 404 \pm 9$ $\mu$m. The discrepancy of this fitted value of $c\tau_{B_s^0}$ from the world average value was included as a systematic uncertainty.

## VIII. AMPLITUDE FIT METHOD

The amplitude fit method [13] is a technique that can be used to calculate an experimental $\Delta m_s$ oscillation limit. This technique requires a modification of Eqs. 10 and 11, yielding the form

$$p_s^{nos/osc}(x, K, d_{pr}) = \frac{K}{c\tau_{B_s^0}} \exp(-\frac{Kx}{c\tau_{B_s^0}}) \cdot 0.5 \cdot (1 \pm \mathcal{D}(d_{pr}) \cos(\Delta m_s \cdot Kx/c) \cdot \mathcal{A}), \tag{24}$$

where $\mathcal{A}$ is now the only fit parameter.

The values of $\Delta m_s$ were changed from 0.5 ps$^{-1}$ to 25 ps$^{-1}$ with a step size of 0.5 ps$^{-1}$. By plotting the fitted value of $\mathcal{A}$ as a function of the input value of $\Delta m_s$, one searches for a peak of $\mathcal{A} = 1$ to obtain a measurement of $\Delta m_s$. For any value of $\Delta m_s$ not equal to the "true" value of $B_s^0$ oscillation frequency, the amplitude $\mathcal{A}$ should be zero. If no peak is found, limits can be set on $\Delta m_s$ using this method. The expected limit (i.e., sensitivity) of a measurement is determined by calculating the probability that at a non-"true"value of $\Delta m_s$ the amplitude could fluctuate to $\mathcal{A}$=1. This occurs at the lowest value of $\Delta m_s$ for which 1.645 $\sigma_{\mathcal{A}} = 1$ for a 95% CL, where $\sigma_{\mathcal{A}}$ is the uncertainty on the value of $\mathcal{A}$ at the point $\Delta m_s$. The limit is determined by calculating the probability that a fitted value of $\mathcal{A}$ could fluctuate to $\mathcal{A} = 1$. This occurs at the lowest value of $\Delta m_s$ for which $\mathcal{A} + 1.645\sigma_{\mathcal{A}} = 1$.

Figure 6 shows the dependence of the parameter $\mathcal{A}$ from Eq. 24, and its uncertainty, on $\Delta m_s$. In the figure, the yellow (light shaded) and green (dark shaded) regions indicate 1.645 times the statistical uncertainty and 1.645 times the statistical plus systematic uncertainties, respectively. A 95% CL limit on the $B_s^0$ oscillation frequency $\Delta m_s > 15.0$ ps$^{-1}$ and expected limit 14.4 ps$^{-1}$ were obtained with statistical uncertainties only.

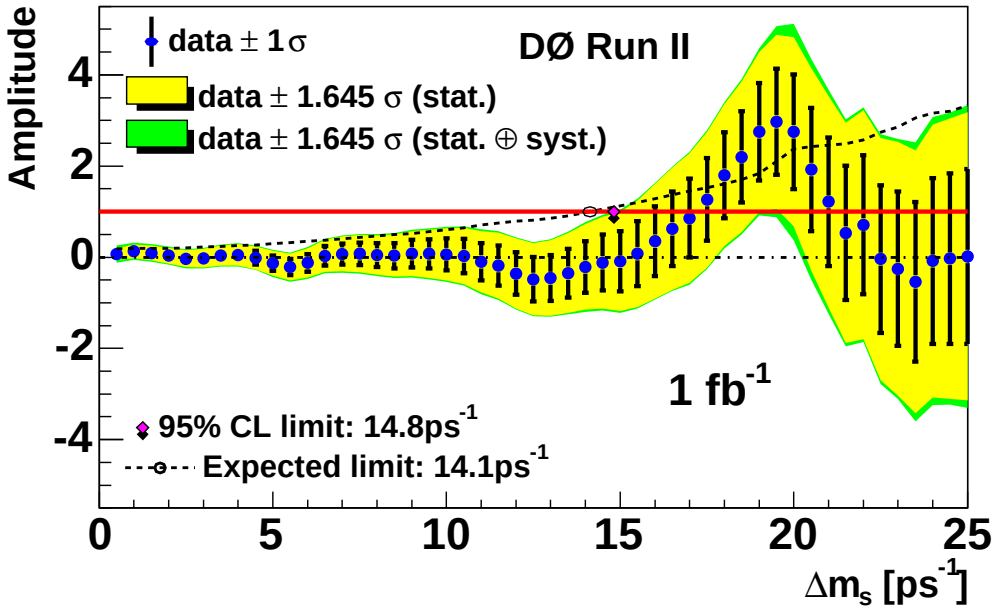


FIG. 6: $B_s^0$ oscillation amplitude with statistical and systematic errors. The red (solid) line shows the $\mathcal{A} = 1$ axis for reference. The dashed line shows the sensitivity including both statistical and systematic uncertainties.

## IX. SYSTEMATIC UNCERTAINTIES AND CROSS-CHECKS

All studied contributions to the systematic uncertainty of the amplitude are listed in Table I. For each $\Delta m_s$ step, the deviations of $\Delta\mathcal{A}$ and $\Delta\sigma_{\mathcal{A}}$ from the default values are given. One can see that the largest deviations come from the uncertainty in the resolution. The resulting systematic uncertainties were obtained using the formula from Ref. [13]

$$\sigma_{\mathcal{A}}^{sys} = \Delta\mathcal{A} + (1 - \mathcal{A})\frac{\Delta\sigma_{\mathcal{A}}}{\sigma_{\mathcal{A}}} \tag{25}$$

and were summed in quadrature. The effect of the systematic uncertainties is represented by the green (dark shaded) region in Fig. 6. Taking into account the systematic uncertainties, we obtained a 95% confidence level limit on the oscillation frequency $\Delta m_s > 14.8$ ps$^{-1}$ and a expected limit of 14.1 ps$^{-1}$.

The decays $B_d^0 \to X\mu^+D^-(\to \phi\pi^-)$ (the left-side peak in Fig. 1) allow for a cross-check of the entire fitting procedure using $B_d^0$ meson decays present in the same data sample as the signal $B_s^0 \to D_s^-\mu^+\nu X$, $D_s^- \to \phi\pi^-$ events. Figure 7 shows the dependence of the parameter $\mathcal{A}$ and its uncertainty on the $B_d^0$ oscillation frequency, $\Delta m_d$, which has a world-average value of $\Delta m_d = 0.502 \pm 0.007$ ps$^{-1}$ [12]. The peak in the amplitude scan at $\Delta m_d \approx 0.5$ ps$^{-1}$ reveals the oscillations in the $B_d^0$—$\bar{B}_d^0$ system. The peak amplitude is in good agreement with unity, which confirms that the dilution calibration is correct.

## X. LIKELIHOOD SCAN

Figure 8(a) shows the dependence of $\mathcal{L}$ on $\Delta m_s$ when the amplitude is fixed to $\mathcal{A} = 1$. The likelihood curve is well behaved and has a preferred value of 19 ps$^{-1}$, with a 90% CL interval from 17 to 21 ps$^{-1}$, assuming Gaussian

TABLE I: Systematic uncertainties on the amplitude. The shifts of both the measured amplitude, $\Delta\mathcal{A}$, and its statistical uncertainty, $\Delta\sigma_{\mathcal{A}}$, are listed. Columns correspond to the different $\Delta m_s$ steps.

| Osc. frequency | | 1 ps$^{-1}$ | 3 ps$^{-1}$ | 5 ps$^{-1}$ | 7 ps$^{-1}$ | 9 ps$^{-1}$ | 11 ps$^{-1}$ | 13 ps$^{-1}$ | 15 ps$^{-1}$ | 17 ps$^{-1}$ | 19 ps$^{-1}$ | 21 ps$^{-1}$ | 23 ps$^{-1}$ | 25 ps$^{-1}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathcal{A}$ | | 0.128 | −0.025 | −0.134 | 0.073 | 0.079 | −0.100 | −0.459 | −0.093 | 0.858 | 2.749 | 1.218 | −0.253 | 0.018 |
| Stat. uncertainty | | 0.090 | 0.124 | 0.167 | 0.231 | 0.299 | 0.410 | 0.504 | 0.659 | 0.864 | 1.068 | 1.413 | 1.690 | 1.920 |
| $\mathrm{Br}(D_sD_s) = 4.7\%$ | $\Delta\mathcal{A}$ | −0.003 | +0.000 | +0.003 | −0.002 | −0.002 | +0.003 | +0.010 | +0.001 | −0.022 | −0.059 | −0.021 | +0.012 | +0.009 |
| | $\Delta\sigma_{\mathcal{A}}$ | −0.002 | −0.003 | −0.004 | −0.005 | −0.006 | −0.009 | −0.010 | −0.014 | −0.018 | −0.023 | −0.029 | −0.035 | −0.040 |
| $\mathrm{Br}(D_s\mu X) = 6.7\%$ | $\Delta\mathcal{A}$ | +0.006 | −0.003 | −0.005 | −0.004 | −0.001 | −0.003 | −0.011 | −0.004 | +0.012 | +0.046 | +0.023 | −0.001 | +0.011 |
| | $\Delta\sigma_{\mathcal{A}}$ | +0.002 | +0.002 | +0.003 | +0.004 | +0.005 | +0.007 | +0.009 | +0.011 | +0.015 | +0.019 | +0.024 | +0.030 | +0.035 |
| $p_{T_\mu} > 6$ GeV/c | $\Delta\mathcal{A}$ | −0.015 | +0.009 | +0.013 | +0.010 | −0.001 | +0.010 | +0.029 | +0.013 | −0.045 | −0.124 | −0.044 | −0.023 | −0.019 |
| | $\Delta\sigma_{\mathcal{A}}$ | −0.004 | −0.006 | −0.008 | −0.011 | −0.014 | −0.019 | −0.024 | −0.031 | −0.042 | −0.054 | −0.066 | −0.081 | −0.093 |
| K-factor variation | $\Delta\mathcal{A}$ | −0.000 | +0.006 | −0.024 | +0.001 | +0.010 | −0.041 | +0.045 | +0.104 | +0.231 | +0.207 | −0.380 | +0.006 | −0.001 |
| 2% | $\Delta\sigma_{\mathcal{A}}$ | +0.000 | +0.001 | +0.002 | +0.004 | +0.007 | +0.012 | +0.011 | +0.027 | +0.025 | +0.059 | +0.040 | +0.049 | +0.050 |
| K-factor distribution | $\Delta\mathcal{A}$ | +0.000 | −0.000 | −0.001 | +0.001 | −0.002 | +0.013 | +0.006 | +0.036 | +0.028 | −0.003 | +0.171 | +0.033 | +0.032 |
| smoothed | $\Delta\sigma_{\mathcal{A}}$ | +0.000 | +0.000 | +0.000 | +0.000 | +0.001 | +0.001 | +0.002 | +0.003 | +0.003 | +0.005 | +0.004 | +0.008 | +0.009 |
| K-factor from | $\Delta\mathcal{A}$ | −0.000 | −0.001 | +0.003 | +0.001 | −0.009 | +0.026 | +0.003 | +0.055 | +0.048 | −0.021 | +0.248 | +0.003 | −0.050 |
| measured momenta | $\Delta\sigma_{\mathcal{A}}$ | +0.000 | +0.000 | +0.000 | +0.001 | +0.001 | +0.002 | +0.003 | +0.005 | +0.004 | +0.006 | +0.006 | +0.005 | +0.011 |
| Fraction of peaking bkg. | $\Delta\mathcal{A}$ | +0.002 | +0.001 | −0.000 | −0.001 | −0.000 | +0.000 | −0.000 | +0.001 | +0.004 | +0.012 | +0.007 | +0.002 | +0.008 |
| (combinatorial bkg.) | $\Delta\sigma_{\mathcal{A}}$ | +0.000 | −0.000 | −0.000 | +0.000 | +0.000 | +0.000 | +0.001 | +0.001 | +0.001 | +0.001 | +0.003 | +0.004 | +0.004 |
| Fraction of peaking bkg. | $\Delta\mathcal{A}$ | +0.001 | −0.000 | −0.002 | −0.000 | −0.002 | −0.007 | −0.016 | −0.013 | +0.004 | +0.055 | +0.014 | −0.035 | −0.021 |
| (signal) | $\Delta\sigma_{\mathcal{A}}$ | +0.001 | +0.001 | +0.001 | +0.002 | +0.002 | +0.004 | +0.005 | +0.007 | +0.012 | +0.014 | +0.026 | +0.034 | +0.039 |
| $c\tau_{B_s}$ | $\Delta\mathcal{A}$ | +0.001 | +0.001 | +0.002 | −0.000 | −0.001 | +0.003 | +0.003 | −0.001 | −0.010 | −0.029 | +0.003 | +0.013 | +0.000 |
| | $\Delta\sigma_{\mathcal{A}}$ | −0.000 | −0.000 | −0.001 | −0.001 | −0.002 | −0.002 | −0.003 | −0.004 | −0.006 | −0.007 | −0.011 | −0.014 | −0.015 |
| uncertainty in | $\Delta\mathcal{A}$ | −0.002 | +0.001 | −0.001 | +0.001 | +0.002 | +0.002 | −0.001 | −0.003 | +0.000 | +0.008 | +0.008 | +0.002 | −0.001 |
| reflection | $\Delta\sigma_{\mathcal{A}}$ | +0.000 | +0.000 | +0.000 | +0.001 | +0.001 | +0.001 | +0.001 | +0.001 | +0.002 | +0.002 | +0.003 | +0.004 | +0.004 |
| Stat. fluctuation of | $\Delta\mathcal{A}$ | −0.001 | +0.000 | +0.000 | +0.001 | −0.000 | −0.001 | −0.000 | +0.003 | +0.008 | +0.016 | +0.011 | +0.004 | +0.009 |
| $N_{D_s}$ | $\Delta\sigma_{\mathcal{A}}$ | +0.000 | +0.001 | +0.001 | +0.001 | +0.001 | +0.002 | +0.002 | +0.003 | +0.004 | +0.004 | +0.008 | +0.009 | +0.009 |
| resolution | $\Delta\mathcal{A}$ | +0.001 | +0.002 | +0.004 | +0.010 | +0.007 | −0.000 | −0.019 | −0.012 | +0.019 | +0.075 | +0.040 | +0.025 | +0.076 |
| (signal) | $\Delta\sigma_{\mathcal{A}}$ | +0.000 | +0.001 | +0.002 | +0.004 | +0.007 | +0.012 | +0.016 | +0.023 | +0.035 | +0.046 | +0.068 | +0.087 | +0.102 |
| resolution | $\Delta\mathcal{A}$ | +0.001 | −0.001 | −0.001 | −0.001 | −0.001 | −0.001 | −0.001 | −0.002 | −0.003 | −0.006 | −0.009 | −0.009 | −0.011 |
| (bkg.) | $\Delta\sigma_{\mathcal{A}}$ | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.001 | −0.001 |
| dilution | $\Delta\mathcal{A}$ | −0.005 | −0.002 | +0.008 | +0.021 | −0.010 | −0.006 | +0.001 | +0.002 | −0.015 | −0.042 | +0.037 | +0.112 | +0.129 |
| | $\Delta\sigma_{\mathcal{A}}$ | −0.001 | −0.001 | −0.001 | −0.002 | −0.003 | −0.005 | −0.004 | −0.005 | −0.004 | −0.002 | −0.017 | −0.018 | −0.018 |
| $\mathcal{F}_{tsens}$ | $\Delta\mathcal{A}$ | −0.010 | +0.006 | +0.003 | +0.003 | +0.003 | +0.000 | −0.002 | −0.004 | −0.005 | −0.004 | −0.000 | +0.001 | −0.005 |
| | $\Delta\sigma_{\mathcal{A}}$ | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | +0.000 | +0.000 | −0.000 |
| $\mathcal{F}_{osc}$ | $\Delta\mathcal{A}$ | −0.005 | −0.000 | +0.000 | −0.001 | −0.001 | −0.001 | −0.001 | −0.002 | −0.002 | −0.003 | −0.004 | −0.006 | −0.006 |
| | $\Delta\sigma_{\mathcal{A}}$ | +0.000 | +0.000 | +0.000 | +0.000 | 0.000 | +0.000 | −0.000 | −0.000 | +0.000 | +0.000 | 0.000 | −0.000 | 0.000 |
| Fit to | $\Delta\mathcal{A}$ | +0.008 | +0.010 | +0.014 | +0.030 | +0.041 | +0.044 | +0.004 | +0.026 | +0.129 | +0.379 | +0.291 | +0.149 | +0.363 |
| VPDL distribution | $\Delta\sigma_{\mathcal{A}}$ | +0.002 | +0.001 | +0.001 | +0.003 | +0.006 | +0.013 | +0.021 | +0.034 | +0.045 | +0.043 | +0.100 | +0.147 | +0.179 |
| Non-zero $\Delta\Gamma$ | $\Delta\mathcal{A}$ | +0.000 | +0.000 | +0.001 | +0.000 | +0.000 | +0.001 | +0.001 | +0.000 | −0.001 | −0.005 | −0.003 | +0.001 | −0.001 |
| | $\Delta\sigma_{\mathcal{A}}$ | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.001 | −0.001 | −0.001 | −0.002 | −0.002 | −0.002 |
| Total syst. | $\sigma_{tot}^{sys}$ | 0.071 | 0.057 | 0.056 | 0.068 | 0.090 | 0.106 | 0.117 | 0.194 | 0.286 | 0.337 | 0.565 | 0.309 | 0.497 |
| Total | $\sigma_{tot}$ | 0.115 | 0.137 | 0.176 | 0.241 | 0.313 | 0.423 | 0.517 | 0.687 | 0.910 | 1.119 | 1.522 | 1.718 | 1.983 |

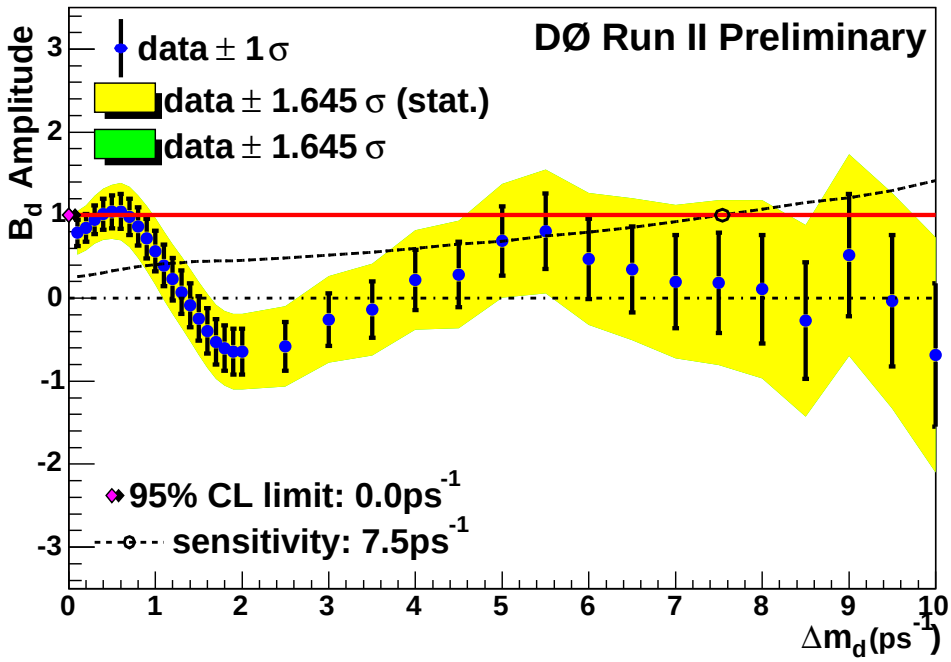


FIG. 7: $B_d^0$ oscillation amplitude with statistical uncertainty only for events in the $D^-$ mass region in Fig. 2. The red (solid) line shows the $\mathcal{A} = 1$ axis for reference. The dashed line shows the expected limit including statistical uncertainties only.

uncertainties. The lower edge of the confidence level interval is close to the world average lower limit, $\Delta m_s = 16.6$ ps$^{-1}$ [1].

To test the statistical significance of the observed minimum in the log-likelihood scan, an ensemble test using the data sample was performed. The sign of the dilution was randomly assigned to each event. This removed any flavor information, simulating a $B_s^0$ oscillation with an infinite frequency. All other information was kept the same as in the standard fit. Many different experiments were performed and the probability of producing a likelihood minimum similar to the one observed in the interval $16 < \Delta m_s < 22$ ps$^{-1}$ was found to be $5.0 \pm 0.3\%$, as shown in Fig. 8(b).

## XI. CONCLUSIONS

Using $B_s^0 \to D_s^- \mu^+ \nu X$ decays, where $D_s^- \to \phi\pi^-$, $\phi \to K^+K^-$, in combination with an opposite-side flavor tagging algorithm and an unbinned fit, we performed a search for $B_s^0 - \bar{B}_s^0$ oscillations.

A log-likelihood scan is well behaved and has a preferred value of $\Delta m_s = 19$ ps$^{-1}$, with a 90% CL interval from 17 to 21 ps$^{-1}$, assuming Gaussian uncertainties. Ensemble tests indicate that if $\Delta m_s$ lies above our region of sensitivity ($> 22$ ps$^{-1}$), only 5.0% of the trial measurements give a measurement similar to or better than our observation anywhere in the window $16 < \Delta m_s < 22$ ps$^{-1}$.

### Acknowledgments

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the DOE and NSF (USA); CEA and CNRS/IN2P3 (France); FASI, Rosatom and RFBR (Russia); CAPES, CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil); DAE and DST (India); Colciencias (Colombia); CONACyT (Mexico); KRF (Korea); CONICET and UBACyT (Argentina); FOM (The Netherlands); PPARC (United Kingdom); MSMT (Czech Republic); CRC Program, CFI, NSERC and WestGrid Project (Canada); BMBF and DFG (Germany); SFI (Ireland); Research Corporation, Alexander von Humboldt Foundation, and the Marie Curie Program.

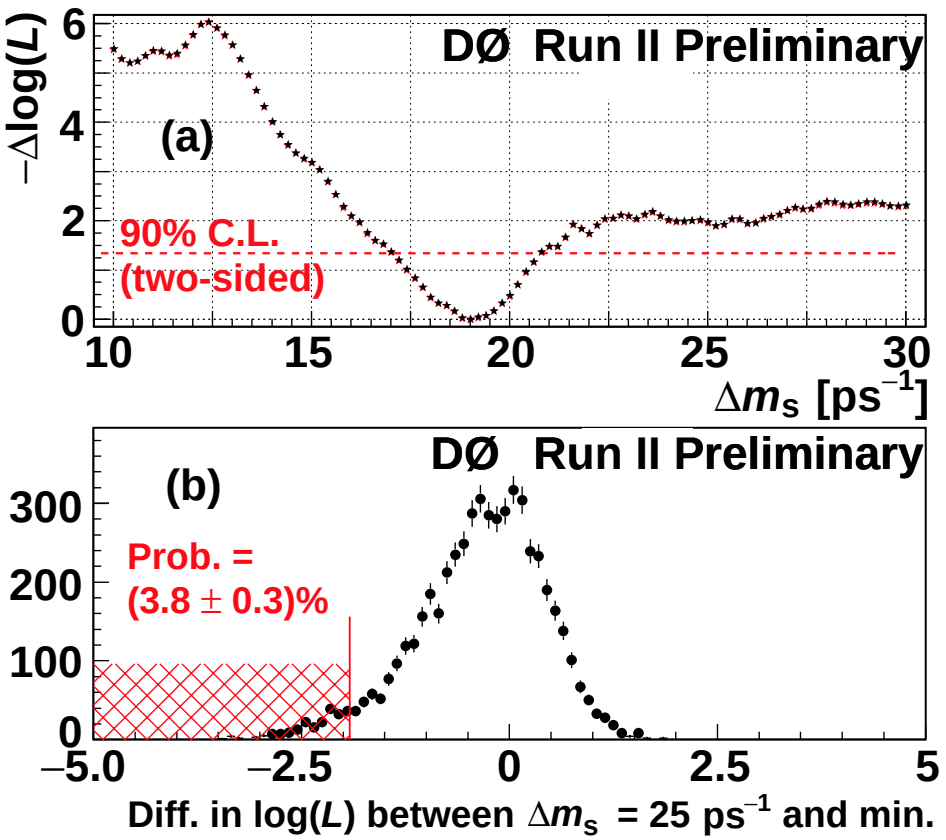


FIG. 8: (a) Dependence of log-likelihood $\mathcal{L}$ on $\Delta m_s$ for $\mathcal{A} = 1$. (b) Difference in $\mathcal{L}$ between $\Delta m_s = 25$ ps$^{-1}$ and a minimum anywhere in the range $16 < \Delta m_s < 22$ ps$^{-1}$ for events with randomly assigned opposite-side tagging information.

## APPENDIX A: VPDL DISTRIBUTION

In Section VII, we presented the value $c\tau_{B_s^0} = 404 \pm 9$ $\mu$m, which was obtained from fitting the VPDL distribution in the total tagged data sample. Figure 9 shows a distribution of the VPDL with optimal fit parameters in the fitting function. Only VPDL *pdf*s were used for this plot.

A systematic uncertainty was assigned to reflect the discrepancies between the data points and fitting function in Fig. 9. Figure 10 shows the same VPDL distribution but with a modified fitting function that better describes the data.


[1] Heavy Flavor Averaging Group, *http://www.slac.stanford.edu/xorg/hfag/index.html*.
[2] CKM Fitter Group, *http://www.slac.stanford.edu/xorg/ckmfitter/ckm_welcome.html*.
[3] UTfit Collaboration, M. Bona *et al.*, J. High Energy Phys. **07**(2005) 028.
[4] V. Abazov *et al.* [DØ Collaboration] *The Upgraded DØ Detector*, submitted to Nucl. Instrum. Methods Phys. Res. A., arXiv:hep-physics/0507191, Fermilab-Pub-05/341-E.
[5] V. M. Abazov *et al.*, Nucl. Instrum. Meth. A **552**, 372 (2005).
[6] S. Catani, Yu.L. Dokshitzer, M. Olsson, G. Turnock, B.R. Webber, Phys.Lett. **B269** (1991) 432.
[7] T. Sjöstrand *et al.*, hep-ph/0108264.
[8] DELPHI Collab., *b-tagging in DELPHI at LEP*, Eur. Phys. J. **C32** (2004), 185-208.
[9] G. Borisov, Nucl. Instrum. Meth. A **417**, 384 (1998).
[10] The DØ Collaboration, *DØ Note 5029*, "$B_d$ Mixing Measurement using Opposite-side Flavor Tagging".
[11] F. James, *MINUIT - Function Minimization and Error Analysis*, CERN Program Library Long Writeup D506, 1998.
[12] S. Eidelman *et al.*, Phys. Lett. **B592**, 1 (2004).
[13] H.G. Moser, A. Roussarie, Nucl.Instrum.Meth.A **384**, 491 (1997).

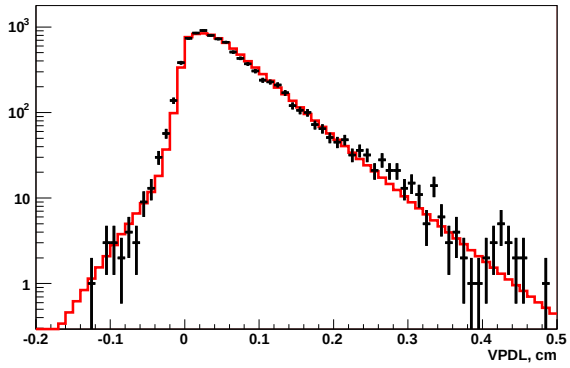


FIG. 9: Distribution of the VPDL in the signal peak region $1.92 < M(D_s^-) < 2.02$ GeV/$c^2$ for all tagged candidates. The points represent the experimental data, the histogram the fitting function.

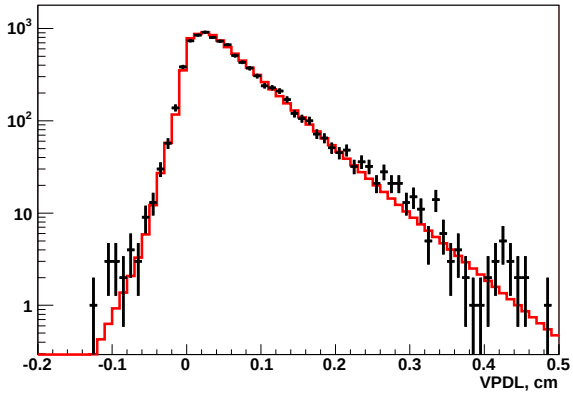


FIG. 10: Distribution of the VPDL in the signal peak region $1.92 < M(D_s^-) < 2.02$ GeV/$c^2$ for all tagged candidates. The points represent the experimental data. The histogram shows an alternate fitting function to that used in Fig. 10.

[14] D.J. Lange, Nucl. Instrum. Meth. A **462**, 152 (2001); for details see *http://www.slac.stanford.edu/˜ lange/EvtGen*.